MODIFIED CARBOHYDRATE PROCESSING ENZYME

Field of Invention

The invention relates to modified carbohydrate processing enzymes and their use in the hydrolysis of glycoside substrates and the synthesis of glycosides.

Background to the Invention

Recent advances in the development of carbohydrate based therapeutics (Koeller and Wong, *Nat. Biotechnol.*, **18** (2000) 835-841), and the limitations of present chemical synthetic methods for producing oligosaccharides, has led to more novel approaches to the synthesis of carbohydrates and their conjugates (Davis, *J. Chem. Soc. Perkin Trans.*, **1** (2000) 2137). One approach to this problem is to carry out such syntheses using carbohydrate processing enzymes such as glycosyltransferases or glycosidases, as a valuable source of catalytic activity for the manipulation of unprotected carbohydrates (Crout and Vic, *Curr. Opin. Chem. Biol.*, **2** (1998) 98-111); Wymer and Toone, *Curr. Opin. Chem. Biol.*, **4** (2000) 110-119; Watt *et al.*, *Curr. Opin. Chem. Biol.*, **7** (1997) 652-660; Kren and Thiem, *Chem. Soc. Rev.*, **26** (1997) 463-473; and Palcic, *Curr. Opin. Biotechnol.*, **10** (1999) 616-624). Glycosidases are simple, robust, soluble enzymes, and in general have been preferred for such glycosynthesis (Scigelova *et al., J. Mol. Catal. B Enzym.*, **6** (1999) 483-494 and Van Rantwijk *et al., J. Mol. Catal. B Enzym.*, **6** (1999) 511-532). Although catalysis of the hydrolysis of glycoside bonds is normally observed, glycosidases may be successfully used to synthesise glycosides through reverse hydrolysis (thermodynamic control) or transglycosylation (kinetic control with activated donors) strategies.

Thus far, improvements in glycosidase synthetic utility have largely focused upon developing new strategies for increasing low product yields (Mackenzie *et al., J. Am. Chem. Soc.*, **120** (1998) 5583-5584), improving regioselectivity of transfer (Prade *et al., Carbohydr. Res.*, **305** (1998) 371-381) or characterising available glycosidases for novel activities (Scigelova *et al., supra*). For example, a major advance in improving yields has been the development of the glycosynthase by Withers and co-workers (Mackenzie *et al., supra*; Mayer *et al., FEBS Lett.*, **466**

(2000) 40-44; Malet and Planas, *FEBS Lett.*, **440** (1998) 208-212; Moracci *et al.*, *Biochemistry* **37** (1998) 17262-17270; Trincone and Perugino, *Bioorg. Med. Chem. Lett.*, **10** (2000) 365-368; Fort *et al., J. Am. Chem. Soc.*, **122** (2000) 5429-5437; and Nashiru *et al., Chem. Int. Ed.*, **40** (2001) 417-420). These nucleophile-less glycosidase mutants are capable of glycosyl transfer in yields of up to 90% using glycosyl fluoride donors, but do not hydrolyse glycoside products and they illustrate well the benefits of glycosidase engineering for creating more synthetically useful catalysts.

Summary of the Invention

The present invention relates to cabohydrate processing enzymes, in particular glycosidase enzymes. Enzymes of the invention are preferably compatible with high temperature and organic solvents, and will form glycosidic linkages between two monosaccharides, without the need for protection or activation steps: a "super β-catalyst".

The inventors have investigated glycoside formation using the retaining β-glycosidase from *Sulfolobus solfataricus* (SsβG). SsβG is thermophilic, and displays tolerance to organic solvents. These attributes highlight the potential of this enzyme as a universal glycosylation catalyst.

The present invention provides a modified polypeptide having carbohydrate processing enzymatic activity, said modification comprising substitution of the amino acid residue forming the catalytic nucleophile of an active site by a less nucleophilic amino acid residue, wherein said less nucleophilic residue retains some nucleophilic activity.

In particular, the invention provides such a polypeptide comprising an amino acid sequence selected from:

(a) the amino acid sequence of SEQ ID NO: 2 comprising substitution of the residue E387 by a less nucleophilic residue;

(b) the amino acid sequence of a family 1 glycosyl hydrolase, comprising a substitution at an amino acid residue equivalent to E387 of SEQ ID NO: 2 by a less nucleophilic residue; and

(c) a variant of (a) or (b) having carbohydrate processing enzymatic activity and comprising a substitution at a position equivalent to E387 of SEQ ID NO: 2 by a less nucleophilic residue,

wherein said less nucleophilic residue retains some nucleophilic activity.

A polypeptide of the invention may further comprise one or more mutations selected to broaden the substrate specificity of the polypeptide compared to a polypeptide not so modified.

The invention also provides polynucleotides encoding polypeptides of the invention, expression vectors comprising such polynucleotides and host cells transformed with such vectors

The invention further provides a method for hydrolysing a β-glycoside, synthesising a β-glycoside or transglycosylation, which method comprises contacting a glycoside substrate with a modified polypeptide of the invention.

Brief Description of the Figures

Figure 1 shows the hydrolysis/transglycosylation process as carried out by glycosidases.

Figure 2 is a scheme showing the formation of glycosides from activated donors (when R = good leaving group).

Figure 3 shows the expression of E387Y SsβG from E. coli strain BL21(DE3). The mutation was effected using the QuikChangeTM strategy. The enzyme was purified by nickel affinity chromatography to yield 28 mhL-1 in > 95 % purity by SDS-PAGE analysis. Lane 1 = loaded protein; lane 4 = wash; lane 5 = eluted E387Y SsβG; lane 6 = SDS-7 markers.

Figure 4 shows the mass spectrometry characterisation of nucleophile trapping by the E387Y mutant.

Figure 5 shows the pH profile of wild type and E387Y SsβG.

Figure 6 shows the transglycosylation activity of the E387Y SsβG mutant. [c] Yields were determined by NMR analysis of the per-acetylated reaction mixture, separated by flash chromatography and based on the recovery of starting material. [d] S = total yield of glycosides/ synthesis products. H = total yield of hydrolysis products.

Figure 7 shows the results of experiments to improve transglycosylation activity. A, b, c, [c] and [d] as in Figure 6. [e] tri = trisaccharides identified by mass spectrometry and anomeric peaks in NMR, not isolated or characterized.

Brief Description of the Sequences

SEQ ID No 1 provides the amino acid sequence of the β-galactosidase of *Sulfolobus solfataricus* as well as the encoding polynucleotide sequence.

SEQ ID No 2 provides the amino acid sequence of the β-galactosidase of *Sulfolobus solfataricus*.

SEQ ID Nos 3 and 4 provide the nucleotide sequence of oligonucleotide primers.

Detailed Description of the Invention

The present invention provides a modified carbohydrate processing enzyme which shows an altered enzymatic activity compared to the unmodified enzyme.

Preferably, a polypeptide suitable for modification is one which has carbohydrate processing enzymatic activity activity prior to modification. For example, Figure 1 shows the routes of hydrolysis and transglycosylation that may be achieved by a glycosidase enzyme. Typically, the modified carbohydrate processing enzyme of the invention will have glycosyl hydrolase, glycosyl synthase and/or transglycosylase activity. The enzyme may possess all three of these activities, any two of them or only one of them. In particular, the enzyme may have transglycosidase synthase activity or may hydrolyse glycoside substrates.

The conditions the enzyme is being used under or the particular concentrations of substrates/products or their ratio may dictate which particular activity an enzyme of the invention displays or which activity predominates at a particular time. In particular, an activated substrate may be used to ensure synthase activity. Alternatively, or additionally, low water activity or sequence modifications may reduce or eliminate hydrolytic activity and allow glycosyl synthase and/or transglycosylase activity to predominate. The conditions and/or concentrations of substrate/products the enzyme of the invention is employed under may be manipulated to ensure that a particular desired activity or activities predominate. For

example the enzyme may be a glycosidase which can hydrolyse glycosidic bonds, but under some conditions can also catalyse their synthesis by transglycosylation.

The modified carbohydrate processing enzymes of the invention are typically produced by modifying a family 1 glycosyl hydrolase. In a preferred embodiment, the family 1 glycosyl hydrolase may be one isolated or originating from a thermophilic organism. For example, the enzyme may be from the thermophilic microbe *Sulfolobus solfataricus* and in particular may be a β-glycosidase from *Sulfolobus solfataricus*. Alternatively, the enzyme to be modified may be another member of the glycosyl hydrolase family 1 such as *Pyrococcus furiosus* β-glucosidase, *Dalbergia cochinchinensis* β-glucoside, *Costus speciosus* β-glycoside hydrolase, human lactase phlorizin hydrolase, myrosinase from *Sinapis alba* or *Staphylococcus aureus* phosphogalactosidase.

The amino acid sequence of β-glycosidase from *Sulfolobus solfataricus* is set out in SEQ ID NO:2. Variants in the sequence of SEQ ID NO: 2 may be present in β-glycosidase obtained from other isolates or strains of *Sulfolobus solfataricus* or other cell types expressing β-glycosidases or enzymes classified as being part of the glycosyl hydrolase family 1. Such variants may be modified in accordance with the invention. Carbohydrate processing enzymes, including family 1 glycosyl hydrolases and in particular β-glycosidases from other *Sulfolobus solfataricus* strains or other cell types expressing such enzymes can be isolated following standard cloning techniques, for example, using the polynucleotide sequence of SEQ ID NO: 1 or a fragment thereof as a probe. The isolated enzymes may then be modified.

The carbohydrate processing enzymes of the invention are modified. These modification(s) have a number of effects on the function and/or activity of the enzyme.

The catalytic nucleophile of a carbohydrate processing enzyme is a nucleophilic amino acid residue that is situated at, or close to, the active site of the enzyme and which acts as a catalyst in the enzymatic reaction controlled by that active site, e.g. by acting as an electron donor. It is known that by replacing this catalytic nucleophile of a glycosyl hydrolase with a non-nucleophilic residue, it is possible to generate an enzyme which lacks hydrolytic activity, but which is still

capable of glycoside synthesis using activated glycosyl donors such as α-glycosyl fluoride. Such mutated enzymes are known as glycosynthases.

According to the present invention, rather than producing a glycosidase that lacks a catalytic nucleophile, the catalytic nucleophile (nucleophilic amino acid residue) is replaced by a less nucleophilic amino acid residue. This allows the enzyme to attain the ability to differentiate between substrates containing good leaving groups and those without. This would allow formation of glycosides from activated donors (see Figure 2; R = good leaving group), but in the transglycosylation products (when R is a poor leaving group), the poor (e.g. tyrosine) nucleophile will be incapable of forming the glycosyl enzyme intermediate, preventing hydrolysis and increasing the transglycosylation yield. The modified enzymes of the invention thus allow one to minimise the hydrolysis of transglycosylation products and consequently greatly improve transglycosylatiojn yields in comparison to wild-type glycosidases.

The unmodified enzyme may accept a number of different substrates. However, the rate of reaction with different substrates may differ significantly. The unmodified enzyme may have higher affinity for a particular substrate, or subgroup of substrates, within the range of possible substrates that it can act on. A modification in accordance with the invention may cause the enzyme to better differentiate between those substrates having good leaving groups and those that do not. A modified enzyme of the invention may thus act preferentially on substrates having good leaving groups over those with poor leaving groups.

A modification in accordance with the invention may increase the activity of the enzyme on one or more substrates which have good leaving groups, whilst having no, or little, effect on the affinity of the enzyme for its other substrates, such as those with poor leaving groups. A modification in accordance with the invention may decrease the activity of the enzyme on one or more substrates which have poor leaving groups, whilst having no, or little, effect on the affinity of the enzyme for substrates with good leaving groups. A modification in accordance with the invention may both increase the activity of the enzyme on one or more substrates which have good leaving groups and decrease the activity of the enzyme on one or more substrates which have poor leaving groups. A good leaving group is generally

the conjugate base of a strong acid. A poor leaving group is generally the conjugate base of a weak acid. The modifications therefore typically lead to an increased differentiation in activity between those substrates having good leaving groups (based on strong acids), and those having poor leaving groups (based on weaker acids). The modified enzyme may thus act preferentially on substrates with particularly good leaving groups. The modification of the invention may therefore introduce a specificity of action of an enzyme when more than one substrate is present in a mixture. For example, an enzyme may be modified such that it will preferentially act on one substrate rather than another when both substrates are present.

Where the modifications of the invention increase the specficity of the enzyme for substrates with good leaving groups, the carbohydrate substrate may comprise an activated donor such as a fluoryl or PNP linked carbohydrate donor. The enzyme may thus catalyse the transfer of the glycoside from the carbohydrate donor onto an acceptor molecule, for example an alcohol acceptor such as, for example another saccharide or polypeptide. In a preferred example, the glycosyl donor used is a β-D-mannoside and it is used to form Man β(1,4) Glc NAc.

Existing glycosynthases may be modified in accordance with the invention to give an enzyme with altered activity. Alternatively, the nucleophilic residue of the active site of a family 1 glycosidase may be mutated at the same time that other modifications are introduced, for example to alter the substrate specificity of the enzyme.

The catalytic nucleophile of the active site, i.e. the amino acid residue which is nucleophilic and acts to catalyse the reaction mediated by the active site, is substituted to generate a modified enzyme of the invention. The identity of the catalytic nucleophile may be identified by methods known in the art, for example by studying the reaction catalysed by the enzyme at a molecular level. Suitable methods for identifying the location of the catalytic nucleophile in a carbohdyrate processing enzyme are described in Okuyama et al (Eur J Biochem 268: 2270-2280, 2001). The present invention is based on the substitution of the amino acid residue that forms the catalytic nucleophile by a less nucleophilic amino acid residue. If the enzyme to be modified contains more than one active site which control more than one enzymatic

activity, then one or more of the active sites may be modified according to the invention by substitution of the catalytic nucleophile of each active site. In the case of a glycosidase enzyme, the catalytic nucleophile to be modified is the amino acid residue that mediates the catalytic activity of hydrolysis and transglycosylation by the enzyme, i.e. residue 387 of SEQ ID NO: 2 in the case of *Sulfolobus solfataricus* β-glycosidase (SsβG).

According to the present invention, the residue that is introduced at the catalytic position in place of the catalytic nucleophile is one exhibiting poorly nucleophilic properties. A nucleophilic residue is one which acts by donating or sharing its electrons, for example aspartic acid or glutamic acid. A poorly nucleophilic residue according to the present invention is one which this nucleophilic ability is weak. For example, the nucleophilic activity of the donor may be weaker than that of glutamic and/or aspartic acid, but some nucleophilic activity may be retained. Preferably a poor nucleophile does have some nucleophilic activity. A poor nucleophile will therefore not be a residue having no nucleophilic activity, i.e. a non-nucleophile such as glycine or alanine. In one embodiment the poor nucleophile is an amino acid residue that is less nucleophilic than the amino acid residue that it replaces, but is not glycine, alanine or serine.. A poor nucleophile is not able to share or donate its electrons to the same extent as a nucleophile, for example because those electrons are drawn away from the donor and towards the rest of the molecule. For example, a poor nucleophile may have a potential electron donor group in which the electrons are stabilised by resonance, or may have an electron withdrawing group attached to the electron donor. In particular, the nucleophilic residue may be substituted with a tyrosine, asparagine, cysteine, glutamine and arginine residue and preferably with a tyrosine residue. The mutations Glu387Tyr, Glu387Asn, Glu387Cys, Glu387Gln and Glu387Arg may be introduced into the sequence of SEQ ID No 2 to generate a glycosynthase or the equivalent mutation may be introduced in other family 1 hydrolases.

Position 387 of SEQ ID NO: 2 is a glutamic acid residue. A "poor" nucleophile in the context of this polypeptide is thus a residue which has less, but still present, nucleophilic activity compared to glutamic acid at this location. An equivalent definition may apply to other, e.g. variant, enzymes of the invention,

where a "poor" nucleophile will be a residue having less, but still present, nucleophilic activity compared to the residue normally present at the nucleophilic location of the active site.

The invention also relates to a variant of SEQ ID NO: 2 having an equivalent modification to those described above. A variant of SEQ ID NO: 2 may be a naturally occurring variant such as one selected from the family 1 of glycosyl hydrolases. A variant may also be a non-naturally occurring variant as described in more detail below. The equivalent amino acid to the residue at position 387 of SEQ ID NO: 2 can be identified by aligning a variant peptide with the sequence of SEQ ID NO: 2. The alignment is selected to provide the best possible match to SEQ ID NO: 2. The equivalent amino acid of any such variant to position 387 may then be identified and modified. Any of the programs discussed herein may be used to perform the alignment and in particular Clustal W based on BLOSUM 42.

The equivalent amino acid residues to residue 387 of SEQ ID No 2 will be a nucleophilic residue, for example glutamic acid or aspartic acid. The equivalent amino acid may also be identified by molecular modelling to identify residues playing the equivalent role to residue 387 of SEQ ID NO: 2. The residue at position 387 of SEQ ID NO: 2 is the nucleophilic residue of the active site of the enzyme. Modelling and active site trapping, as well as sequence alignment, may be used to identify the active site nucleophile which may then be mutated in accordance with the invention.

A variant polypeptide having an amino acid sequence which varies from that of SEQ ID NO: 2 may be modified in accordance with the present invention. A variant for use in accordance with the invention is one having carbohydrate processing enzymatic activity. The variant may be, or may be derived from, any family 1 glycosyl hydrolase. A modified variant in accordance with the invention is preferably one which demonstrates a reduced hydrolysis of transglycosylation products and an improved ability to differentiate between substrates having good and poor leaving groups, compared to a variant sequence not so modified.

In some cases the enzyme may recognise and act on the same substrates as the unmodified enzyme, but have a different substrate affinity for each substrate.

A variant of SEQ ID NO: 2 may be a naturally occurring variant which is expressed by another strain of *Sulfolobus solfataricus* or other cell type. Such variants may be identified by looking for carbohydrate processing enzymatic activity in those cells which have a sequence which is highly conserved compared to SEQ ID NO: 2. Such proteins may be identified by analysis of the polynucleotide encoding such a protein isolated from an alternative strain, for example, by carrying out the polymerase chain reaction using primers derived from portions of SEQ ID NO: 2 or degenerate primers based on evolutionarily conserved regions of SEQ ID NO: 2.

Variants of SEQ ID NO: 2 include sequences which vary from SEQ ID NO: 2 but are not necessarily naturally occurring carbohydrate processing enzymes. Over the entire length of the amino acid sequence of SEQ ID NO: 2, a variant will preferably be at least 30% homologous to that sequence based on amino acid identity. The variant may, for example, be at least 40% homologous, more preferably be at least 50% homologous and still more preferably be more than 65% homologous to the amino acid sequence of SEQ ID NO: 2. In some embodiments the polypeptide will be at least 75% homologous, preferably at least 80% homologous and even more preferably the polypeptide is at least 85% homologous to SEQ ID NO: 2. The polypeptide may be at least 90% homologous and still more preferably be at least 95%, 97% or 99% homologous to the amino acid sequence of SEQ ID NO: 2. A variant may be a variant of any family 1 glycosyl hydrolase with one of the percentages of sequence homology specified above.

These percentages of homology may, for example, be over at least 30 amino acids, preferably over at least 40 amino acids and even more preferably over 50 amino acids. The percentages of homology may be over at least 75 amino acids, preferably at least 100, more preferably over 150 amino acids and in some cases will be over the entire length of the variant. In some cases they may be over all but 10, preferably all but 20, more preferably all but 30 and even more preferably all but 50 contiguous amino acids of the variant. There may be at least 80%, for example at least 85%, 90% or 95%, amino acid identity over a stretch of 40 or more, for example 60, 100 or 120 or more, contiguous amino acids ("hard homology").

In a preferred embodiment of the invention the variant will comprise a region which has one of the levels of amino acid sequence homology specified herein to the

region(s) of the amino acid sequence of the polypeptide that forms the active site. The variant may be one of any family 1 hydrolase as long as the residue equivalent to 387 of SEQ ID NO: 2 is modified according to the present invention.

Preferably sequence alignment and the determination of homology may be performed using ClustalW based on a BLOSUM42 matrix.

Amino acid substitutions may be made to the amino acid sequence of SEQ ID NO: 2, for example from 1, 2 or 3 to 10, 20 or 30 substitutions. Such modifications may be introduced into any family 1 glycosyl hydrolase. Conservative substitutions may be made, for example, according to the following table. Amino acids in the same block in the second column and preferably in the same line in the third column may be substituted for each other:

ALIPHATIC	Non-polar	G A P
		I L V
	Polar – uncharged	S T M
		N Q
	Polar – charged	D E
		K R
AROMATIC		H F W Y

One or more amino acid residues of the amino acid sequence of SEQ ID NO: 2 may alternatively or additionally be deleted. From 1, 2 or 3 to 10, 20 or 30 residues may be deleted, or more. Polypeptides of the invention also include fragments of the above-mentioned sequences. Such fragments retain carbohydrate processing enzymatic activity. Fragments may be at least from 10, 12, 15 or 20 to 60, preferably100 or 200, 300 or more amino acids in length.

Such fragments may be used to produce chimeric enzymes using portions of enzyme derived from other carbohydrate processing enzymes such as, for example, glycosidases.

One or more amino acids may be alternatively or additionally added to the polypeptides described above. An extension may be provided at the N-terminus or C-terminus of the amino acid sequence of SEQ ID NO: 2 or polypeptide variant or

fragment thereof. The, or each, extension may be quite short, for example from 1 to 10 amino acids in length. Alternatively, the extension may be longer. A carrier protein may be fused to an amino acid sequence according to the invention. A fusion protein incorporating the polypeptides described above can thus be provided.

Polypeptides of the invention may be in a substantially isolated form. It will be understood that the polypeptide may be mixed with carriers or diluents which will not interfere with the intended purpose of the polypeptide and still be regarded as substantially isolated. A polypeptide of the invention may also be in a substantially purified form, in which case it will generally comprise the polypeptide in a preparation in which more than 90%, e.g. 95%, 98% or 99%, by weight of the polypeptide in the preparation is a polypeptide of the invention.

Polypeptides of the invention may be modified for example by the addition of histidine residues to assist their identification or purification or by the addition of a signal sequence to promote their secretion from a cell where the polypeptide does not naturally contain such a sequence. It may be desirable to provide the polypeptides in a form suitable for attachment to a solid support. For example the polypeptides of the invention may be modified by the addition of a cysteine residue.

A polypeptide of the invention above may be labelled with a revealing label. The revealing label may be any suitable label which allows the polypeptide to be detected. Suitable labels include radioisotopes, e.g. ^{125}I, ^{35}S, enzymes, antibodies, polynucleotides and linkers such as biotin. Labelled polypeptides of the invention may be used in diagnostic procedures such as immunoassays in order to determine the amount of a polypeptide of the invention in a sample.

The proteins and peptides of the invention may be made synthetically or by recombinant means. The amino acid sequence of proteins and polypeptides of the invention may be modified to include non-naturally occurring amino acids or to increase the stability of the compound. When the proteins or peptides are produced by synthetic means, such amino acids may be introduced during production. The proteins or peptides may also be modified following either synthetic or recombinant production.

The proteins or peptides of the invention may also be produced using D-amino acids. In such cases the amino acids will be linked in reverse sequence in the

C to N orientation. This is conventional in the art for producing such proteins or peptides.

A number of side chain modifications are known in the art and may be made to the side chains of the proteins or peptides of the present invention. Such modifications include, for example, modifications of amino acids by reductive alkylation by reaction with an aldehyde followed by reduction with $NaBH_4$, amidination with methylacetimidate or acylation with acetic anhydride.

The polypeptides of the invention may be introduced into a cell by *in situ* expression of the polypeptide from a recombinant expression vector. The vector may be stably integrated into the genome of the cell. The expression vector optionally carries an inducible promoter to control the expression of the polypeptide.

Such cell culture systems in which polypeptides of the invention are expressed may be used in assay systems.

A polypeptide of the invention can be produced in large scale following purification by high pressure liquid chromatography (HPLC) or other techniques after recombinant expression as described below.

The enzymes of the present invention are modified. By this it is meant that one or more amino acid sequence changes have been introduced into the enzyme in comparison to the unmodified sequence of the protein. Thus, typically a wild type enzyme will have had amino acid sequence changes introduced to produce the modified enzyme. The amino acid sequence changes introduced will affect the nucleophilic residue of the active site of the enzyme, for example amino acid position 387 of SEQ ID NO: 2 or the equivalent residues of other family 1 glycosyl hydrolases. The unmodified form of the enzyme will typically be the naturally occurring form of the enzyme. However, the amino acid substitutions of the invention may also be introduced into mutant and variant forms of family 1 glycosyl hydrolases.

In a preferred embodiment of the invention the enzyme is a modified form of β-galactosidase of *Sulfolobus solfataricus*, β-galactosidase of *Sulfolobus shibatae*, β-galactosidase of *Sulfolobus acidocaldarius*, β-galactosidase of *Thermoplasma volcanium*, β-galactosidase of *Pyrococcus furiosus*, β-glycosidase of *Agrobacterium tumefaciens*, β-D-glucoside glucohydrolase of *Bacillus circulans*, β-D-glucoside

glucohydrolase of *Agrobacterium sp.*, β-glucoside of *Rhizobium meliloti*, β-D-glucoside of *Bacillus halodurans*, β-D-glucoside glucohydrolase of *Paenibacillus polymyxa*, β-galactosidase glucohydrolase of *Pyrococcus woesi*, β-glucoside of *Dalbergia cochinchinensis*, Furostanol β- glucoside of *Costus specious*, Lactase phlorizin hydrolase of *Homo sapiens*, Myrosinase of *Sinapis alba*, or 6-phospho-beta-galactosidase of *Staphylcoccus aureus* which comprises one or more of the modifications of the invention. A modified polypeptide of the invention may comprise a variant of such sequences.

A modified polypeptide in accordance with the present invention may additionally comprise one or more further modifications from a naturally occurring or other known sequence. For example, any combination of the modifications described herein may be present. For example, an enzyme in accordance with the present invention may be further modified to alter its substrate specificity..

In one aspect, a polypeptide according to the present invention further incorporates one or more mutations in the region of the active site. Such an enzyme may further include a mutation in one or more of the amino acid residues 432 (glutamine), 433 (tryptophan) or 439 (methionine) of SEQ ID NO: 2. Alternatively the enzyme of the invention may be a family 1 glycosyl hydrolase comprising at least one mutation at an amino acid residue equivalent to W433, E432 or M439 of SEQ ID NO:2. The invention also encompasses variants of these sequences. Such mutants are described in more detail in Corbett *et al* (FEBS Letters (2001) 509: 355-360), which also describes how such mutants can be obtained and how the equivalent positions in other enzymes can be derived.

The mutation will typically be an amino acid substitution of W433, E432 or M439 or of the equivalent residues in other family 1 glycosyl hydrolases. Alternatively, the mutation may be a deletion comprising one or more of these residues or an insertion or duplication affecting these residues. Preferred modifications include mutation of the glutamine, tryptophan or methionine residues or their equivalents to cysteine. Replacement with other amino acids is also contemplated. For example, the residues may be replaced by alanine or valine. In cases where more than one amino acid substitution is made the amino acids introduced may be the same or different at some or all of the sites substituted. For

example, the amino acids at positions 432,433 and 439 may all be replaced with cysteine or with any combination of cysteine, alanine and/or valine.

The invention also relates to a variant of SEQ ID NO: 2 having an equivalent modification to those described above. An equivalent position to these residues may be determined as described above and in Corbett *et al* (supra).

The change in substrate specificity caused by a particular mutation may relate to any or all of the activities of the enzyme. For example, it may relate to the hydrolase, synthase and/or transglycosylase activites of the enzyme and in particular to the hydrolase or synthase activities of the enzyme.

The K_M for a particular substrate may be, for example, increased due to the introduction of the modification(s) of the invention by a factor of from 1.1 to 50 fold, preferably by a factor of from 3 to 40 fold, more preferably by a factor of from 5 to 25 fold and even more preferably by a factor of from 10 to 15 fold. This may be accompanied by reduction in K_{CAT} by a factor of from 1.1 to 50 fold, preferably by a factor of from 3 to 40 fold, more preferably by a factor of from 5 to 25 fold and even more preferably by a factor of from 10 to 15 fold for the same substrate. The value of K_{CAT} may be increased, for example, by a factor of from 1.1 to 250, preferably by a factor of from 2 to 200, more preferably by a factor of from 5 to 150, even more preferably by a factor of from 10 to 100 and still more preferably by a factor of from 20 to 75. These changes will typically be seen for a natural substrate of the enzyme and in particular for any of glucoside (Glc), galactoside (Gal), fucoside (Fuc), xyloside (Xyl) mannoside (Man) and/or glucuronide (GlcA) substrates. In particular, the changes will be seen with glucoside, galactoside, fucoside and/or mannoside substrates and preferably with glucoside and/or galactoside substrates..

The substrate specificity of an enzyme in accordance with the invention can be monitored *in vitro* or *in vivo*, for example in accordance with the methods described in more detail below. In particular, assays can be carried out to monitor activity of the enzyme on particular substrates and in particular glycosidase substrates. Suitable substrates include glucosides, galactosides, fucosides, β-mannosides and β-glucuronides.

The assay may measure glycoside synthesis, hydrolysis and/or transglycosylation. Activity may be assayed using a chromophore such as, for

example, paranitrophenol (PNP). The chromophore may be conjugated to a sugar as the carbohydrate donor molecule in glycoside synthesis or transglycosylation or as a substrate for hydrolysis. The release of the chromophore may be monitored to follow the course of the reaction and hence determine the activity of the enzyme. The release of leaving groups such as the fluoride ion, when a glycosyl fluoride is employed as a carbohydrate donor, may also be monitored to determine enzyme activity. The release of the fluoride ions may be measured using a fluoride electrode. Enzyme activity may also be monitored by using mass spectroscopy to monitor the formation of the product ion or decrease in the amount of the substrate ion.

The invention also relates to polynucleotides encoding the modified carbohydrate processing enzymes. A polynucleotide of the invention typically is a contiguous sequence of nucleotides which is capable of hybridising selectively with the coding sequence of SEQ ID NO: 1 or to the sequence complementary to that coding sequence. Polynucleotides of the invention include variants of the coding sequence of SEQ ID NO: 1 which encode the amino acid sequence of SEQ ID NO: 2. Such polynucleotides additionally incorporate one or more modification to encode a modified polypeptide as described in more detail above.

A polynucleotide for use in the invention and the coding sequence of SEQ ID NO: 1 can typically hybridize at a level significantly above background or alternatively the complement of such a sequence can. Background hybridization may occur, for example, because of other cDNAs present in a cDNA library. The signal level generated by the interaction between a polynucleotide of the invention and the coding sequence of SEQ ID NO: 1 is typically at least 10 fold, preferably at least 100 fold, as intense as interactions between other polynucleotides and the coding sequence of SEQ ID NO: 1. The intensity of interaction may be measured, for example, by radiolabelling the probe, e.g. with ^{32}P. Selective hybridization is typically achieved using conditions of medium to high stringency (for example 0.03M sodium chloride and 0.003M sodium citrate at from about 50°C to about 60°C).

A nucleotide sequence capable of selectively hybridizing to the DNA coding sequence of SEQ ID NO: 1 or to the sequence complementary to that coding sequence will be generally be at least 30%, preferably at least 40% and even more

preferably at least 50% homology to the coding sequence of SEQ ID No. 1. Sequence homology corresponds to sequence identity. In some embodiments it will be at least 60%, preferably at least 70% and more preferably at least 80%, homologous to the coding sequence of SEQ ID NO: 1 or its complement over a region of at least 20, preferably at least 30, for instance at least 40, 60 or 100 or more contiguous nucleotides or, indeed, over the full length of the coding sequence. Thus there may be at least 85%, at least 90% or at least 95% nucleotide identity over such regions.

Any combination of the above mentioned degrees of homology and minimum size may be used to define polynucleotides of the invention, with the more stringent combinations (i.e. higher homology over longer lengths) being preferred. Thus for example a polynucleotide which is at least 85% homologous over 25, preferably over 30, nucleotides forms one aspect of the invention, as does a polynucleotide which is at least 90% homologous over 40 nucleotides.

Nucleotide homology may be determined using various BLAST programs and in particular PSI-BLAST. Polynucleotide variants for use in the invention may be identified by performing PSI-BLAST searches of SWISSPROT and TREMBL to a family 1 glycosyl hydrolase, including any of those mentioned herein, and in particular to the amino acid sequence of SEQ ID No. 1.

Alternatively, the UWGCG Package provides the BESTFIT program which can be used to calculate homology (for example used on its default settings) (Devereux *et al* (1984) *Nucleic Acids Research* **12**, p387-395). The PILEUP and BLAST algorithms can be used to calculate homology or line up sequences (such as identifying equivalent or corresponding sequences (typically on their default settings), for example as described in Altschul S. F. (1993) J Mol Evol 36:290-300; Altschul, S, F *et al* (1990) J Mol Biol 215:403-10.

Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pair (HSPs) by identifying short words of length W in the query sequence that either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighbourhood word score threshold

(Altschul *et al*, supra). These initial neighbourhood word hits act as seeds for initiating searches to find HSP's containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Extensions for the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T and X determine the sensitivity and speed of the alignment. The BLAST program uses as defaults a word length (W) of 11, the BLOSUM62 scoring matrix (see Henikoff and Henikoff (1992) *Proc. Natl. Acad. Sci.* USA 89: 10915-10919) alignments (B) of 50, expectation (E) of 10, M=5, N=4, and a comparison of both strands.

The BLAST algorithm performs a statistical analysis of the similarity between two sequences; see e.g., Karlin and Altschul (1993) *Proc. Natl. Acad. Sci.* USA 90: 5873-5787. One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a sequence is considered similar to another sequence if the smallest sum probability in comparison of the first sequence to the second sequence is less than about 1, preferably less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

Polynucleotides of the invention may comprise DNA or RNA. They may also be polynucleotides which include within them synthetic or modified nucleotides. A number of different types of modification to polynucleotides are known in the art. These include methylphosphate and phosphorothioate backbones, addition of acridine or polylysine chains at the 3' and/or 5' ends of the molecule. For the purposes of the present invention, it is to be understood that the polynucleotides described herein may be modified by any method available in the art. The invention also includes protein nucleic acid (PNA) molecules comprising the sequences of the invention.

Polynucleotides of the invention may be used to produce a primer, e.g a PCR primer, a primer for an alternative amplification reaction, a probe e.g. labelled with a

revealing label by conventional means using radioactive or non-radioactive labels, or the polynucleotides may be cloned into vectors. Such primers, probes and other fragments will be at least 15, preferably at least 20, for example at least 25, 30 or 40 nucleotides in length, and are also encompassed by the term polynucleotides of the invention as used herein. The invention also provides a microarray comprising such polynucleotides.

Polynucleotides such as a DNA polynucleotide and primers according to the invention may be produced recombinantly, synthetically, or by any means available to those of skill in the art. They may also be cloned by standard techniques. The polynucleotides are typically provided in isolated and/or purified form.

In general, primers will be produced by synthetic means, involving a step wise manufacture of the desired nucleic acid sequence one nucleotide at a time. Techniques for accomplishing this using automated techniques are readily available in the art.

Longer polynucleotides will generally be produced using recombinant means, for example using PCR (polymerase chain reaction) cloning techniques. This will involve making a pair of primers (e.g. of about 15-30 nucleotides) to a region of the gene which it is desired to clone, bringing the primers into contact with DNA obtained from a suitable cell, performing a polymerase chain reaction under conditions which bring about amplification of the desired region, isolating the amplified fragment (e.g. by purifying the reaction mixture on an agarose gel) and recovering the amplified DNA. The primers may be designed to contain suitable restriction enzyme recognition sites so that the amplified DNA can be cloned into a suitable cloning vector.

Although in general the techniques mentioned herein are well known in the art, reference may be made in particular to Sambrook *et al*, 1989.

Polynucleotides or primers of the invention may carry a revealing label. Suitable labels include radioisotopes such as ^{32}P or ^{35}S, enzyme labels, or other protein labels such as biotin. Such labels may be added to polynucleotides or primers of the invention and may be detected using techniques known *per se*.

Polynucleotides of the invention can be incorporated into a recombinant replicable vector. The vector may be used to replicate the nucleic acid in a

compatible host cell. Thus in a further embodiment, the invention provides a method of making polynucleotides of the invention by introducing a polynucleotide of the invention into a replicable vector, introducing the vector into a compatible host cell, and growing the host cell under conditions which bring about replication of the vector. The vector may be recovered from the host cell. Suitable host cells are described below in connection with expression vectors.

Preferably, a polynucleotide of the invention in a vector is operably linked to a control sequence which is capable of providing for the expression of the coding sequence by the host cell, i.e. the vector is an expression vector. Such expression vectors can be used to express the polypeptide of the invention.

The term "operably linked" refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences. Multiple copies of the same or different modified carbohydrate processing enzyme genes may be introduced into the vector.

Such vectors may be transformed into a suitable host cell to provide for expression of a polypeptide of the invention. Thus, a polypeptide according to the invention can be obtained by cultivating a host cell transformed or transfected with an expression vector as described above under conditions to provide for expression of the polypeptide, and recovering the expressed polypeptide.

The vectors may be for example, plasmid, virus or phage vectors provided with an origin of replication, optionally a promoter for the expression of the said polynucleotide and optionally a regulator of the promoter. The vector may be an artificial chromosome such as a human or yeast artificial chromosome. The vectors may contain one or more selectable marker genes, for example a tetracycline resistance gene. Promoters and other expression regulation signals may be selected to be compatible with the host cell for which the expression vector is designed. Multiple copies of the same or different modified glycosidase gene in a single expression vector, or more than one expression vector each including a modified glycosidase gene which may be the same or different may be transformed into the host cell.

Host cells transformed (or transfected) with the polynucleotides or vectors for the replication and expression of polynucleotides of the invention will be chosen to be compatible with the said vector. In one embodiment of the invention lypholised host cells are produced and used directly as biocatalysts.

The present invention also provides non-human animals comprising a polynucleotide encoding a modified enzyme of the invention. The non-human transgenic animal may, for example, be a rodent, such as a mouse or rat, or an animal such as a pig, sheep or cow. The invention also provides a plant comprising a polynucleotide encoding a modified polypeptide of the invention.

Where the amino acid at position 433, 432 or 439 of SEQ ID NO: 2 or an equivalent position is substituted by cysteine, the cysteine may be chemically modified so as to change the substrate specificity of the enzyme. The cysteine may be modified so as to comprise a positively-charged group, a negatively-charged group or an uncharged group. The positively charged group may be of formula -(CH_2)n-N^+R_3, wherein n is a positive integer from 1 to 4 and each R, which may be the same or different, is H or a C_1-C_4 alkyl group (preferably a methyl group). A preferred positively charged group is -$CH_2CH_2NMe_3^+$. The negatively-charged group may be of formula -(CH_2)n-SO_3^- or -(CH_2)n-COO^-, wherein n is a positive integer from 1 to 4. Preferably, the negatively-charged group is -CH_2CH_2-SO_3^-. The uncharged group may be a C_1-C_4 alkyl group and preferably is methyl.

An enzyme in accordance with the invention can be used *in vitro*, for example, bound to an immobile substrate. The enzyme can be immobilised through the addition of a binding sequence such as a His-tag or maltose binding site or by using a general immobiliser. The immobilised enzyme can then be used in the conversions described herein.

The activity of a modified enzyme in accordance with the invention may be monitored by carrying out assays *in vitro* or *in vivo*, that is within a host cell, to monitor for carbohydrate processing activity of the enzyme. Such assays may include monitoring for the production of glycosides.

The modified enzymes in accordance with the present invention can be used in any methods involving glycosyl synthase, transglycosylase and/or hydrolase activity using glycoside substrates. They can be used wherever it is desired to a form

β glycoside bond. The enzymes may be used in methods in which one or more glycoside substrates, such as one or more glucoside, galactoside, fucoside, mannoside or glucuronide substrates are incubated together with the modified enzyme. Preferably, the glycoside is β-mannoside. Preferably, in accordance with present invention more than one substrate is provided in the same reaction vessel to yield a library of different glycosides. Such substrates may include a natural substrate of the unmodified polypeptide and one or more non-natural substrates, that is substrates that are not usually accepted by the unmodified polypeptide. In a particularly preferred aspect of the present invention the enzymes may be used to differentiate between substrates present in a mixture where the desired substrates contain good leaving groups. Alternatively, reactions may be run in parallel using the enzyme of the invention where the only change between reactions is that a different substrate is employed and hence a different glycoside produced. Such reactions may be run in multiwell plates to allow for the individual screening of each glycoside produced in a high throughput assay. The enzymes may also be used more generally to improve yields, for example by reducing hydrolysis of transglycosylation products the transglycosylation yield may be improved.

The enzymes of the invention may be used in glycoside synthesis and in transglycosylation, they may also be employed in glycoside hydrolysis. Using the enzymes practically any β glycoside linkage may be synthesised or alternatively hydrolysed.

The enzymes of the invention may be used to generate an array of molecules conjugated to carbohydrates. They may be used to generate glycoproteins and in particular O-linked glycosylations, where typically the sugar group is conjugated to a serine or a threonine residue. The enzymes may be used to help produce recombinant proteins which have the same or similar glycosylations to naturally occurring versions of the proteins. The enzymes may be used to generate antibiotics and in particular macrolide antibiotics. They may be used in the food industry, for example to achieve depulping. They may also be used in detergents.

The enzymes may be used in therapy both as therapeutic molecules themselves and in the generation of therapeutic molecules. Thus the enzymes may

be used in the treatment of a human or animal subject. The enzymes may be used in methods of treatment of the human or animal body by surgery or therapy.

The enzymes may be used to develop glycoconjugates for use in LEAP (lectin enzyme activated prodrug system). Lectins are found on the surface of cells. There are a variety of different lectins with certain ones only being found on a specific cell type or on specific groups of cell types. In LEAP glycoconjugates comprising a carbohydrate group capable of binding a specific lectin and an enzyme capable of activating a prodrug are generated and administered to a subject to which the prodrug is also given. The lectin binding group of the conjugate targets it to the specific cell type or types expressing the target lectin and hence the prodrug is only activated at the surface of the specific cell types. Thus LEAP allows drugs to be targeted to a specific class of cells through the lectins that they express and this can be used for a variety of functions including eliminating undesired cells. LEAP is described in WO 02/080980 which is incorporated herein by reference in its entirety. The enzymes of the invention can be employed in the production of any of the glycoconjugates described in WO 02/080980.

In glycoside synthesis using the enzyme of the invention the molecule glycosylated may be a saccharide or a different molecule such as a polypeptide. Multiple glycosylations of the same molecule may occur and, for example, di-, tri-, tetra or oligosaccharides may be generated. These may be generated, for example, by multiple step-wise glycosyl additions or by addition of an oligosaccharide to the target molecule. Branched oliosaccharides may also be added to a target molecule using the enzyme of the invention.

Examples

Creating the mutants

The gene encoding the thermophilic, retaining, exo-β-glycosidase, from *Sulfolobus solfataricus* (SSβG, EC 3.2.1.23), was originally isolated and sequenced from the *Sulfolobus solfataricus* strain MT4 (Cubellis *et al., Gene* (1990) 94, 89-94) and is classified as a member of the glycosyl hydrolase family 1 (Henrissat, (1991) *Biochem J.*, 280, 309-316). This robust, thermophilic enzyme is ideal (Pisani *et al.,*

Eur. J. Biochem. **187** (1990) 321-328; Moracci *et al., Protein Eng.*, **9** (1996) 1191-1195; and Nucci *et al., Biotechnol. Appl. Biochem.*, **17** (1993) 239-250). It can be routinely expressed in *Escherichia coli* (Moracci *et al., Enzym. Microb. Technol.*, **17** (1995) 992-997). Its 3D structure has a classic $(\alpha/\beta)_8$ TIM barrel (Banner *et al., Nature* **255** (1975) 609-614) containing a radial active site channel in a kink of the 5th α/β repeat (Aguilar *et al., J. Mol. Biol.*, **271** (1997) 789-802). The nucleophilic residue of the active site of this enzyme is located at position 387 (glutamic acid). Substrate specificity in this enzyme is associated with two residues in the binding site, glutamate 432 and methionine 439.

Reagents, enzymes and bacterial strains

The wild type sequence, *lac S,* encoding the β-glycosidase from *Sulfolobus solfataricus* (SsβG), was amplified by PCR from *Sulfolobus* genomic DNA, using the following primers:

5': CCATGGGACACCACCACCACCACCACTCATTAC (SEQ ID No.3)

3': CTCGAGTTAGTGCCTTTATGGCTTTACTGGAGGTAC (SEQ ID No.4)

The 5' primer introduced an N-terminal *Nco* I site and a 6 x His tag immediately following the ATG initiation codon. The 3' primer introduced a *Xho* I site after the stop codon. The PCR product was cloned into pCR2.1 (Invitrogen) and individual clones were sequenced to verify that no errors had been introduced.

Electrocompetent *Escherichia coli* strain BL21(DE3) and His-bind Nickel resin were obtained from Novagen. 4-Methylumbelliferyl-β-D-glycoside substrates were purchased from Sigma. *Pfu*-turbo DNA polymerase was obtained from Stratagene and *Nco* I, *Xho* I restriction endonucleases, T4 DNA ligase from Promega, UK. Oligonucletoide primers were obtained from MWG BioTech GmBH and Cruachem Ltd. DNA sequencing was carried out by the DNA Sequencing Service, Dept. Biological Sciences, Durham, using standard protocols on Applied Biosystems DNA Sequencers.

Construction, selection and screening of the single point mutants

Mutations were introduced into the *lac S* gene coding sequence (in pCR2.1) according to the Stratagene QuickChange mutagenesis system, using the suppliers'

protocols. Individual point mutations were verified by DNA sequence analysis. Wild type and mutated coding sequences were cloned into the *Nco* I / *Xho* I sites of expression vector pET-24-d(+) (Novagen) and transformed into *E. coli* BL21(DE3). Putative transformants were identified by colony PCR using the SSβG coding sequence primers. Selected clones were checked by DNA sequencing to confirm the mutation, and the absence of unintended PCR-introduced base changes.

Overexpression and purification of the His_6-tagged mutant enzymes

Selected clones were grown in LB medium containing kanamycin (50 μg/ml), at 37°C to an O.D. of 0.6 at 600 nm, and the target were proteins induced by the addition of 0.1M IPTG. Cells were harvested by centrifugation, resuspended in $1/10^{th}$ volume of column loading buffer (5mM imidazole, 20mM Tris, 0.5M NaCl, pH 7.8), and lysed using a Soniprep 150 Sonicator. The suspension was recentrifuged to pellet cell debris (10000 rpm, 30 min), and the His_6-tagged recombinant proteins were purified from the supernatant using Ni-chelation chromatography (wash buffer, 60mM imidazole, 20mM Tris, 0.5M NaCl, pH 7.8; elution buffer 300mM imidazole, 20mM Tris, 0.5M NaCl, pH 7.8). The eluted protein peak was dialysed against 50mM sodium phosphate buffer, (pH 6.5), and stored at 4°C. Protein concentration was quantified by the method of Bradford 1976 *Anar Biochem.*, 151, 196-204 (reagents from Biorad, Netherlands). Purified proteins were analysed by SDS-polyacrylamide gel electrophoresis, gel fitration chromatography and ESMS (Micromass LCT, ± 8Da). The E387Y SsβG mutant yielded 28 mgL^{-1} in > 95% purity (see Figure 3).

Characterisation of the kinetic properties of enzymes

Determination of the Michaelis-Menten parameters for wild type and mutant enzymes was performed at pH 6.5 at 80°C for a range of substrates, which allowed activities to be determined with a high degree of sensitivity (Tables below).

Parameters were determined by the method of initial rates. Activity of wild type, E432C, W433C and M439C mutants was measured in time course assays of the hydrolysis of 4-methylumbelliferyl-β-D-glycosides (β-D-gluco, β-D-galacto, β-D-fuco, β-D-manno, β-D-xylo, β-D-glucurono) at 5-15 concentrations (0.001-1.5 mM)

incubated at 80°C in 50 mM sodium phosphate buffer, pH 6.5. Reactions were terminated at 2, 5, 10, 15 min by the addition of 100µl of ice cold 1M Na_2CO_3, pH 10 and analyzed (Labsystems Fluoroscan Ascent plate reader, excitation 460 nm, emission 355 nm). K_M and k_{cat} were derived by fitting the initial rates to hyperbolic Michaelis-Menten curves using GraFit 4 (Erithacus Software Ltd, Staines, UK).

A similar method was used to determine the Michaelis Menten parameters for the E387Y mutant.

Under these optimised assay conditions, the glucoside(Glc), galactoside (Gal) and fucoside (Fuc) substrates were hydrolysed well by the wild type enzyme, but the xyloside (Xyl) substrate was hydrolysed relatively poorly (approx. 3% of turnover as determined by k_{cat} compared with β-D-glucoside).

The hydrolysis of pNPβGal (*p*-nitrophenyl β-D-galactoside) and pNPβGlc (*p*-nitrophenyl β-D-glucoside) by the E387Y mutant was greatly decreased compared to that by the wild-type enzyme. pNPβGal and pNPβGlc parameters were measured by following *p*-nitrophenol release at 405 nm. Methyl β-D-galactopyranoside (MeβGal) parameters were measured by 1H NMR.

Substrate	Enzyme, SSβG-	K_m, mM	k_{cat}, s^{-1}	k_{cat} / K_m, $s^{-1}mM^{-1}$
4-MUGlc	WT	0.046 ± 0.017	140 ± 20	2900
	E432C	0.34 ± 0.07	5.1 ± 0.5	15
	W433C	1.61 ± 0.35	33 ± 5	20
	M439C	0.068 ± 0.028	190 ± 40	2900
4-MUGal	WT	0.066 ± 0.017	98 ± 7	1490
	E432C	0.47 ± 0.14	5.4 ± 0.8	11
	W433C	2.2 ± 1.2	14 ± 6	6.3
	M439C	0.083 ± 0.016	94 ± 11	1130
4-MUFuc	WT	0.011 ± 0.002	80 ± 2	7300
	E432C	0.34 ± 0.04	18 ± 1	53
	W433C	0.41 ± 0.09	31 ± 3	76
	M439C	0.023 ± 0.005	91 ± 8	4000
4-MUMan	WT	0.036 ± 0.009	1.8 ± 0.2	50

	E432C	0.90 ± 0.26	2.8 ± 0.7	3.2
	W433C	0.18 ± 0.02	0.92 ± 0.05	5.1
	M439C	0.042 ± 0.015	2.3 ± 0.4	53
4-MUXyl	WT	0.13 ± 0.03	3.8 ± 0.3	30
	E432C	1.26 ± 0.21	2.8 ± 0.3	2.2
	W433C	0.59 ± 0.19	1.5 ± 0.3	2.5
	M439C	0.068 ± 0.007	9.3 ± 0.2	136
4-MUGlcA	WT	1.3 ± 0.4	0.81 ± 0.18	0.60
	E432C	NAD[a]	NAD	NAD
	W433C	NAD	NAD	NAD
	M439C	1.4 ± 0.6	1.3 ± 0.4	0.92

Substrate	Enzyme, SSβG-	K_m, mM	k_{cat}, s^{-1}	k_{cat} / K_m, $s^{-1}mM^{-1}$
pNPβGal	WT	0.46	5.07	11140
	E387Y	0.17	7.59 x 10-3	44.4
pNPβGlc	WT	0.20	3.47	17777
	E387Y	0.16	1.79 x 10-3	14.9
MeβGal	WT		None*	
	E387Y		None*	

* no activity was measured after incubation with MeβGal (10 mM) for 1 day.

Mechanistic studies

The E387Y mutant enzyme was characterised by nucleophile trapping. The results of the nucleophile trapping experiment were analysed by mass spectrometry analysis (Figure 4). The formation of a trapped glycosyl-enzyme intermediate with DNPFG (mass +165) corresponded to loss of activity.

The pH profile of the wild type and E387Y mutant SsβG enzymes were analysed. A pH profile for the E387Y mutant enzyme as compared to the wild type SsβG enzyme was obatined (Figure 5). In the E387Y mutant, the basic leg is shifted

up by 0.6 pK_a units. This implies alteration of the general acid pKa, e.g. position E206 that has not been mutated. This may indicate a reverse protonation mechanism pH profile, as described in Joshi *et al* (J. Mol Biol (2000) 299:255).

Transglycosylation activity

The E387Y SsβG mutant enzyme was assayed for transglycosylation of a variety of substrates in 50mM phosphate buffer at pH 6.5, 45/80°C (see Figure 6). Yields were determined by NMR analysis of the per-acetylated reaction mixture, separated by flash chromatography and based on the recovery of starting material.

The results indicate that aromatic sugar donors were preferred, which may be due to stacking interactions in more than 1 subsite. 1-6 and 1-3 regioselectivity and β-only stereoselectivity were also observed. The reaction times at 80°C were shorter, corresponding to lower hydrolysis yields. The enzyme showed broad acceptor specificity, processing galacto-, manno- and gluco- acceptors. Wild type SsβG gave no transglycosylation products under identical conditions.

Conditions were varied to optimise the transglycosylation activity of the enzyme (see Figure 7). Increasing the acceptor concentration increased glycoside synthesis. Increasing enzyme concentration increased hydrolysis. A higher reaction concentration slightly improved conversion but did not affect yields. The transglycosylation yields seen were a >90% improvement over unmodified glycosidases (~50%).